TL;DR Hazard: A Comprehensive Study of Levelsquatting Scams

Kun Du[1], Hao Yang[1], Zhou Li[2], Haixin Duan[3(✉)], Shuang Hao[4], Baojun Liu[1], Yuxiao Ye[1,5], Mingxuan Liu[1], Xiaodong Su[6], Guang Liu[7], Zhifeng Geng[8], Zaifeng Zhang[9], and Jinjin Liang[9]

[1] Tsinghua University, China
{dk15,yang-h16,lbj15,liumx18}@mails.tsinghua.edu.cn
[2] University of California, Irvine zhou.li@uci.edu
[3] Tsinghua University, Beijing National Research Center for Information Science and Technology, China duanhx@tsinghua.edu.cn
[4] University of Texas at Dallas shao@utdallas.edu
[5] yeyuxiao@outlook.com
[6] suxiaodong.sxd@gmail.com
[7] lg2001607@163.com
[8] zhifeng.geng@qq.com
[9] Network security Research Lab at Qihoo 360, China
{zhangzaifeng,liangjinjin}@360.cn

Abstract. In this paper, we present a large-scale analysis about an emerging new type of domain-name fraud, which we call *levelsquatting*. Unlike existing frauds that impersonate well-known brand names (like google.com) by using similar second-level domain names, adversaries here embed brand name in the subdomain section, deceiving users especially mobile users who do not pay attention to the entire domain names.

First, we develop a detection system, LDS, based on passive DNS data and webpage content. Using LDS, we successfully detect 817,681 levelsquatting domains. Second, we perform detailed characterization on levelsquatting scams. Existing blacklists are less effective against levelsquatting domains, with only around 4% of domains reported by VirusTotal and PhishTank respectively. In particular, we find a number of levelsquatting domains impersonate well-known search engines. So far, Baidu security team has acknowledged our findings and removed these domains from its search result. Finally, we analyze how levelsquatting domain names are displayed in different browsers. We find 2 mobile browsers (Firefox and UC) and 1 desktop browser (Internet Explorer) that can confuse users when showing levelsquatting domain names in the address bar.

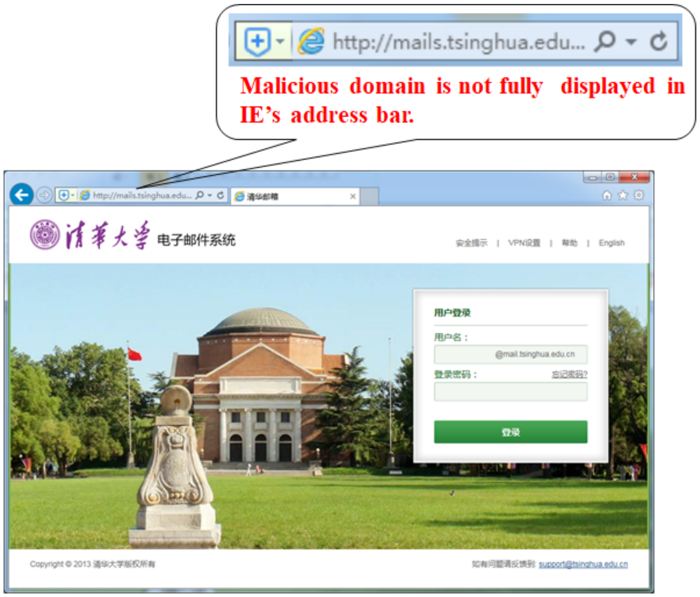
In summary, our study sheds light to the emerging levelsquatting fraud and we believe new approaches are needed to mitigate this type of fraud.

Keywords: LDS, DNS, levelsquatting

1 Introduction

Fast-paced reading is favored in the Internet age. Lengthy articles are less likely to be read and often receive comments like *TL;DR* (short for *Too long; didn't read*) [1]. While impatience to long text may leave valuable information overlooked, negligence to a long domain name can lead to much worse consequences.

As a real-world example, Figure 1 shows a phishing website with a long domain name, `mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com`, displayed in IE browser's address bar with default settings. The domain name is so lengthy that only the subdomain `mails.tsinghua.edu.cn` can be displayed, which is identical to the authentic login domain name of Tsinghua university. A user can be deceived to put her login credential when visiting this website.


http://mails.tsinghua.edu...
清华邮箱
清华大学 电子邮件系统
安全提示 | VPN设置 | 帮助 | English
用户登录
用户名：
@mail.tsinghua.edu.cn
登录密码：
忘记密码?
登录
Copyright © 2013 清华大学版权所有
如有问题请反馈到: support@tsinghua.edu.cn

Malicious domain is not fully displayed in IE's address bar.

Fig. 1: An example of Levelsquatting domain displayed in IE.

We term this type of fraud as *levelsquatting*. Adversaries here create domains by using its *subdomain* section to impersonate a *brand domain*. Levelsquatting scams bring cybercriminals several benefits: (1) This type of attack is more deceptive (compared to traditional domain squatting), since the displayed part of the domain name can have quite legitimate looking in both desktop and mobile browsers; (2) Adversaries can create subdomains to impersonate arbitrary brand domains. If they use e2LDs(effective second level domain names) for the same purpose, they have to find ones not registered yet. and (3) Adversaries can leverage mechanisms of name servers that controlled by themselves, like wildcard DNS, to manage a large pool of levelsquatting concurrently. In this work, we perform the first large-scale analysis to understand this type of fraud.

Finding levelsquatting domains. To discover levelsquatting domains, we have developed a system named called `LDS` (Levelsquatting Detection System), which monitors large volume of *passive DNS data* and identifies levelsquatting. `LDS` first searches for the levelsquatting candidates by matching a list of popular domain names. Then for each candidate, it collects WHOIS information, page content, visual appearance, and performs a three-stage detection procedure. After sampling and manually verification, we confirm LDS can work effectively. As described in Section 3, LDS achieves the precision of 96.9% on a sample of our dataset.

Discoveries. The amount of levelsquatting domains discovered by LDS is 817,681, which enable us to conduct a comprehensive study of levelsquatting scams. We highlight our findings below.

(1) We find a new type of attack that impersonates search engines. For example, the domain www.baidu.com.baidu-service.com has identical appearance as Baidu and it can even returns meaningful search results when being queried. The goal of adversaries here is to insert illegal ads, *e.g.*, gamble promotions, in the returned results. In total, we find 13,331 fake search-engine websites. We report them to Baidu security team, and all of them have been confirmed malicious.

(2) While a levelsquatting domain can be created by adding a subdomain record into the DNS zone file, we find *wildcard DNS record* is used more often for management ease: 517,839 (63.33%) levelsquatting FQDNs (fully qualified domain names referring to absolute domain names) or 41,389 (64.55%) e2LDs have wildcard DNS records.

(3) The effectiveness of blacklists regarding levelsquatting is very limited. We check the identified levelsquatting domains on PhishTank [10] and VirusTotal [11]. Only around 4% of the them have been captured by VirusTotal and PhishTank respectively.

(4) We conjecture that the rise of levelsquatting attack is attributed to the problematic design of modern browsers. In fact, we investigate and show that some mobile browsers (*e.g.*, Firefox and UC) and desktop browsers (*e.g.*, Internet Explorer 9 on Windows 7) fail to display levelsquatting FQDNs correctly, making users vulnerable to this fraud. As a result, we suggest these browser manufacturers to adjust their UI and highlight the e2LD section.

In summary, our work makes the following contributions.

(1) We perform the first large-scale study of levelsquatting fraud using a detection system LDS we developed.

(2) We make an in-depth measurement study of the identified levelsquatting domains.

(3) We check levelsquatting on PC and mobile browsers and find several visual issues that can confuse users. We suggest browser manufactures to fix those issues and highlight the e2LD section more clearly.

2 Background

In this section, we first give a brief overview of existing methods for subdomain creation. Then we define levelsquatting and describe the scope of this study. Finally, we survey existing attacks against brand names that have been extensively studied and compare them with levelsquatting.

Subdomain creation. In this work, we consider a domain name as *FQDN*, its right part offered by registrar (*e.g.*, GoDaddy [12]) as *e2LD* and its left part as *subdomain*. To learn whether a domain is managed by a registrar, we check if it is one level under an effective

[10] https://www.phishtank.com/

[11] https://www.virustotal.com/

[12] https://www.godaddy.com/

top-level domain (eTLD) (*e.g.*, `.com` and `.co.uk`) [13], an approach commonly used by existing works [4].

There are three types of DNS records that can create subdomain, `A`, `AAAA` and `CNAME` records. The first two associate a subdomain with an IP v4/v6 address, *e.g.*, `<b.example.com A 93.184.216.34>`. `CNAME` specifies the alias of another canonical domain, *e.g.*, `<www.example.com CNAME example.com>`. Additionally, the owner can specify a *wildcard* record, by filling the subdomain part with a character `*`, which will capture DNS requests to any subdomain not specified in the zone file.

Levelsquatting. A registrar usually enforces no extra restriction on subdomain creation, if the whole domain name complies with the IETF standard [5]. Such loose policy unfortunately allows attackers to create a subdomain impersonating a well-known brand without any hurdle. We name such fraud domains as *levelsquatting* domains. More concretely, it contains a well-known brand (*e.g.*, `google.com`) in its *subdomain* section, while the e2LD section does not belong to the brand owner.

Whether a domain is created for levelsquatting depends on its similarity to a known brand in both its subdomain and e2LD sections. For the subdomain section, we assume attackers: (1) use the exact brand name without any typo (*e.g.*, `go0gle.com.example.com` is excluded); (2) keep the entire e2LD section of the targeted brand within the subdomain section (*e.g.*, `google.example.com` is excluded); (3) target a brand's FQDN as well in addition to its e2LD (*e.g.*, `accounts.google.com.example.com` is included). We choose these criteria to reduce the computation overhead (*e.g.*, finding all brand typos is computationally expensive) while achieve good coverage.

Comparison to domain-squatting. Previous studies have revealed many tricks adopted by adversaries to impersonate a brand. *Domain-squatting* is arguably the most popular approach. In this approach, adversaries *buy an e2LD* that looks similar to a brand domain and fool users who cannot distinguish the difference. This can be done through typo-squatting [6], bit-squatting [7], homophone-squatting [8], homograph-squatting [9] and *etc*. A recent work by Kintis *et al.* covers combo-squatting, in which case attackers combine brand name with one or more phrases (*e.g.*, `youtube-live.com`) and register the e2LD [10]. Despite the high similarity, these approaches will fail if the user is careful enough when reviewing the domain name.

However, a recent attack called *punycode scam* takes one step further to erase the visual difference. Punycode is a way to represent a Unicode letter using ASCII character set. But many Unicode letters look almost the same as ASCII letters (*e.g.*, Cyrillic "a" and Latin "a"). They can be abused to construct scam domains looking exactly the same as brand domains [11, 12].

All approaches listed above require attackers to *buy* e2LDs *similar* to the targeted brand. The monetary cost is still non-negligible and the choices are usually limited. In comparison, creating levelsquatting domain needs virtually *zero* cost and the choices are *unlimited*. Moreover, when the domain is displayed in a defective browser, discerning the difference is much more difficult.

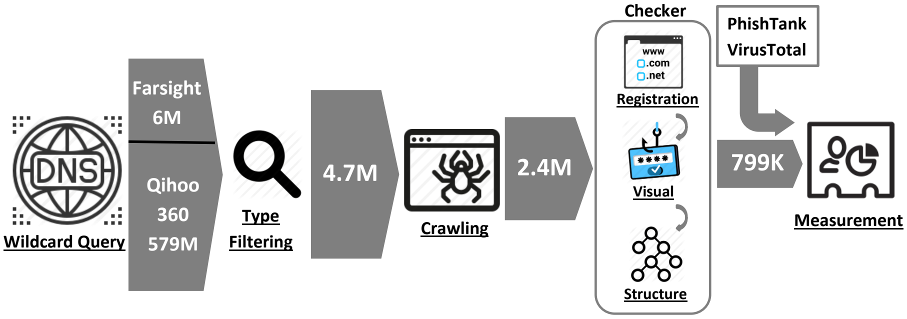


Fig. 2: Processing flow of `LDS`. The number in the figure refers to the number of records remained after each filtering step.

3 Finding levelsquatting Domains

While levelsquatting domains are spotted in wild occasionally [2], there is no systematic study measuring the scale and characterizing the purpose. A large volume of samples is essential to yield meaningful insights into this phenomenon, but so far the coverage from public sources is still limited (see Section 5.2 for more details). To overcome the issue of data scarcity, we build an system named `LDS` (Levelsquatting Detection System) to automatically discover scam levelsquatting domains. At high level, `LDS` selects candidate domains from passive DNS data and identifies scam ones based on the combination of registration-, structural- and visual-analysis. Below we first give an overview of `LDS` and then dive into the details of each component.

3.1 System Overview

The top challenge we need to address here is how to discover a large amount of levelsquatting domains efficiently. Although some registrars (*e.g.*, VeriSign) have published zone files they managed, subdomains are not included. Whether a subdomain exists can be learned through issuing DNS query, but enumerating all subdomains is impossible. Our solution, on the other hand, is to examine the domain resolutions logged by passive DNS collectors. We scan two passive DNS datasets offered by Farsight [14] and Qihoo 360 [15] in this research.

Brand selection. Although any brand may be subjected to levelsquatting attack, impersonating well-known brands accords with the best interest of attackers. In this study, we select e2LD from Alexa top 10K list [16] (named Dom_{Alexa}) for detection. This dataset yields a decent coverage of web categories (46 categories labeled by Alexa [17] are included). Next, we construct a list of wildcard strings (*e.g.*, `*.google.com.*`) and submit them to passive DNS service. In the end, we obtain a corpus of 586,197,541 DNS logs. We filter logs matching `A`, `AAAA` and `CNAME` in record type and extract domain names. We collect 4,735,289 domains as candidates (named Dom_{All}).

[13] We use the public suffix list provided by `https://publicsuffix.org/` to match eTLD.
[14] `https://www.dnsdb.info/`
[15] `https://www.passivedns.cn/`
[16] `http://s3.amazonaws.com/alexa-static/top-1m.csv.zip`
[17] `https://www.alexa.com/topsites/category`

Design and data collection. Through an initial exploration on a small subset within Dom_{All}, we gain three insights about levelsquatting domains. First, many of them have been leveraged to deliver phishing content with similar visual appearances to the targeted brand domains [13]. Second, attackers prefer to use off-the-shelf website template to reduce development cost [14, 15], introducing irregular similarity among pages of levelsquatting domains. Third, registration information of levelsquatting e2LD and brand e2LD are usually irrelevant. Motivated by these insights, we build a crawler infrastructure to query WHOIS information from registrars, download homepage and capture screenshots for each domain in Dom_{All}.

We obtain 2,473,809 valid pages from Dom_{All} and we label this set as Dom_{Sus}. We notice that almost half of Dom_{All} become expired during our research. This is because adversaries here prefer to e2LD with short lifetime to reduce their cost, illustrated by previous work [15]. Every domain in Dom_{Sus} is examined by a detection component based on registration-, structural-, and visual-features and the alarmed domain is considered as levelsquatting (the set is named Dom_{LD}). Figure 2 illustrates the processing flow and the implementation details are elaborated in the following chapter.

3.2 Implementation of Checkers

We develop three checkers to exam each domain in Dom_{Sus}. All these three checkers are sequential. At the high level, a domain is labeled *suspicious* if registration information mismatches correspondent brand domain in Dom_{Alexa}. Structural and visual representation check similarity between Dom_{Sus} or Dom_{Alexa}. We consider a domain as levelsquatting if two checkers alarm. The details of each checker is elaborated below.

Registration checker. We query public WHOIS servers to obtain registration information for e2LDs in Dom_{Sus} and Dom_{Alexa}. Though a levelsquatting domain can pretend by manipulating the subdomain section, faking registration information is not always feasible. In fact, not all the WHOIS fields can be controlled by attackers, *e.g.*, register email and registration date. Although adversaries can utilize "Domain Privacy Protection" service to hide their tracks, they cannot rely on brand domain use the same service.

From WHOIS servers, we obtain 58,372 and 10,000 valid records for e2LDs in Dom_{Sus} and Dom_{Alexa} [18]. For every WHOIS record associated with Dom_{Sus}, we extract email address, telephone number, creation date, expiration date, and match them with Dom_{Alexa}. The domains having zero overlap will be further inspected by the structural- and visual- checker.

Structural checker. As the second step, we inspect the homepage under each domain. On one hand, malicious pages tend to share the same structure due to the use of web templates. On the other hand, when a malicious page is designed for phishing, its structure should resemble to the brand domain. As a result, we compare each page structural similarity in Dom_{Sus} and Dom_{Alexa} by using "Page Compare library" [19].

Visual checker. In this step, we aim to determine whether the levelsquatting domain runs a phishing page mimicking one in Dom_{Alexa}. We look into the visual similarity between them. As the first step, our crawler launches a browser instance and visit homepages

[18] We are not able to obtain WHOIS records for all e2LDs within Dom_{Sus} because they have become expired when we queried.

[19] `https://github.com/TeamHG-Memex/page-compare`

in Dom_{Alexa} and Dom_{Sus} by using selenium library [20]. We take a screen shot for each domain. Then we check structural similarity between each image in Dom_{Sus} and Dom_{Alexa} by using skimage [21].

By using both structural and visual checkers, we can filter out non-malicious levelsquatting domains. Similar to our approach, DeltaPhish [19], also exploits the structural and visual similarity to detect phishing pages. Though DeltaPhish extracted more features, it relies on a pre-labeled training dataset and the computation is more time-consuming. Our approach is training-free and more efficient.

4 Evaluation

The precision of LDS. LDS detects 817,681 levelsquatting FQDNs (Dom_{LD}) and we want to learn how accurate the result is. In the beginning, we use "query" mode of VirusTotal API [22] to get URL report for every detected levelsquatting FQDN and use the number of alarms to determine whether it is scam. But it turns out that most of the domains are not even been submitted to VirusTotal (more details in Section 5.2). Therefore, we have to resort to manual verification. However, manually confirming all of them within a reasonable time is impossible. As an alternative, we sample FQDNs randomly and validate them for 10 rounds. We calculate precision rate for each round and consider the average value as the true precision rate.

In each round, we first sample 1,000 results and check whether the FQDN is used for phishing, *e.g.*, stealing login credentials. For the remaining ones, our validation rules focus on the strategies adopted by attackers. In particular, we first compare two pages crawled by common browser user-agent and spider user-agent strings, determining if cloaking performed, which is widely used for Blackhat SEO. Then we follow the method proposed by Wang *et al.* [17] to find cloaking pages: if there is no similarity in visual effect or page structures between two pages, the domain is labeled as cloaking. Next, we go through the page content and check if it is used to promote illegal business like porn, gamble or fake shops. We also examine e2LD's WHOIS information and consider it a true positive when the domain is recently registered by a non-authoritative party. After 10 rounds calculation, we get the system precision rate is 96.9%.

Analysis of false positives. We conservatively treat the false positives rate 3.1%. But a close look suggests none of them is absolutely innocent. Among these 310 domains, 178 of them show regional news, but none of their sources are well known and the same content/page structure are found, which indicate they might serve spun content for spam purposes [18]. The other 132 domains all display a message showing that the domain is expired. However, when we revisited them one month later, 118 of them showed more than 2 ads about lottery and porn. We speculate these domains might be purchased later by attackers or just use expired pages occasionally to avoid detection.

5 Measurement

In this section, we present our analysis about levelsquatting domains. We first describe the dataset we use. Then, we evaluate how effective the current defense stands against

[20] https://www.seleniumhq.org/

[21] https://scikit-image.org/

[22] The "query" mode retrieves the prior scanning result of a URL that *has been submitted* to VirusTotal by another user.

levelsquatting and how popular levelsquatting is used for scam activities. Next we examine the statistics of the lexical features, including the popularity of different prefixes in subdomains. Finally, we take a deep look into the infrastructure behind levelsquatting domains.

5.1 Datasets

To enrich the diversity of the levelsquatting domains, in addition to the 799,893 domains captured by LDS, we also acquire data from PhishTank and VirusTotal. The summary is listed in Table 1.

PhishTank (DS_{PT}). Levelsquatting is supposed to be used a lot for phishing attacks. As a result, we download all URLs submitted to PhishTank between May 2016 to July 2017, with 1,025,336 records in total, and search for levelsquatting FQDNs. We use the same check algorithm described in Section 3 and get 14,387 levelsquatting FQDNs in the end.

VirusTotal (DS_{VT}). Another data source is VirusTotal, a well-known public service offering URL and file scanning. We download the feed from February to April, 2017, accounting for 160,399,466 URLs in total. After filtering, we obtain 3,528 levelsquatting FQDNs (all of them are alarmed by *at least two* blacklists).

Combining the three datasets, we obtain 817,681 unique levelsquatting FQDNs (we name the entire set DS_{All}), mapped to 64,124 e2LDs. The overlap of the three datasets is small: only 127 FQDNs or 40 e2LDs from DS_{LDS} are also contained in DS_{PT} and DS_{VT}.

Table 1: Summary of datasets.

Notation	Source	Period	# FQDNs	# e2LDs
DS_{LDS}	LDS	03.2017-04.2017	799,893	58,988
DS_{PT}	PhishTank	05.2016-07.2017	14,387	3,887
DS_{VT}	VirusTotal	02.2017-04.2017	3,528	1,289
$DS_{Overlapped}$	—	—	127	40
Sum (DS_{All})	—	—	817,681	64,124

5.2 Impact of Levelsquatting

Blacklist is a common first-line defense against malicious URLs, but according to our study, its coverage on levelsquatting domains is quite limited. Our conclusion comes from a coverage test on VirusTotal: we queried all 817,681 FQDNs from DS_{LDS} using VirusTotal API under "query" mode, and found only 39,249 are alarmed, accounting for 4.80% of DS_{LDS}. It turns out that most of the domains (618,374, 75.63%) are not even submitted to VirusTotal.

Although levelsquatting has been observed in the wild as an attack vector for phishing, whether it has become a popular option for the phishing purpose is unclear yet. The answers seems negative: 332,007 distinct FQDNs (covering 1,025,336 URLs) are obtained from PhishTank but DS_{PT} only has 14,387 (4.33% of 332,007) FQDNs. As another supporting evidence, most of the domains recorded by PhishTank are short, each of which in average consists of only 2.83 levels.

Prefix. Attackers are free to add prefixes in front of a brand, in order to impersonate a specific brand domain. To learn their preference, we have extracted all prefixes and

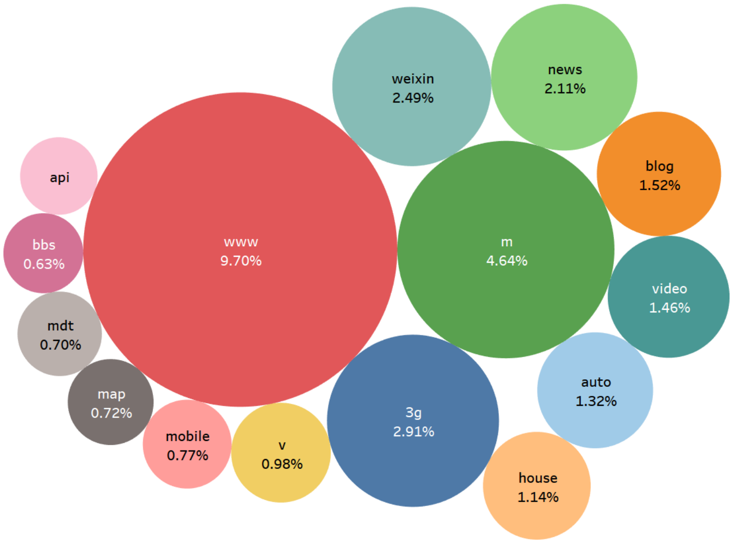


Fig. 3: Top 15 prefix keywords.

counted the number of appearance among DS_{All}. Top 15 prefixes with their occupied percentage are shown in Figure 3. Among them, `www.` is chosen most frequently (79,338 or 9.70% of DS_{All}). The top 15 prefixes show up 31.09% of all levelsquatting domains. Prefix known to be associated with mobile services, like `m.`, `3g.` and `weixin.` (representing WeChat, the top mobile chat app in China), are ranked highly, suggesting that attackers actively exploit the display vulnerabilities in mobile devices (discussed in Section 7).

5.3 Infrastructure

Levelsquatting domains serve as the gateway to attackers' infrastructure. For better understanding, we first look into the IP addresses and registrants behind, then we analyze domains with wildcard DNS record, distribution in new gTLD and HTTPS certificates they deployed.

IP addresses. We performed DNS queries on all levelsquatting FQDNs in DS_{All} to obtain their IP addresses by using pydig [23]. In total, 710,347 (86.87%) requests returned valid results and 54,118 IPs were obtained. We show the top 10 IP addresses that levelsquatting domains prefer in Table 2. From this table we can see that the top 10 servers host more than 38% of total levelsquatting domains.

Registrants. We are interested in who actually control the levelsquatting domains. Hence we select WHOIS records of domains in DS_{All} and obtain 58,372 valid records in total. By grouping the domains with registrant email addresses, we find that 23.41% of them are under 10 email addresses. We list these registrants in Table 3. We search email addresses for relevant information, find that many of them belong to professional domain brokers who own massive amount of domains. Similar observations were also described in previous works looking into the underground economy [3] and blackhat SEO [15].

[23] `https://github.com/shuque/pydig`

Table 2: Top 10 IP addresses of malicious levelsquatting domains.

No.	IP	ASN	Location	Count of levelsquatting FQDNs	Percentage
1	69.172.201.153	AS19324	US	76,387	9.34%
2	185.53.179.8	AS61969	Europe	48,932	5.98%
3	199.59.242.150	AS395082	US	35,327	4.32%
4	202.181.24.196	AS55933	Australia	34,395	4.21%
5	205.178.189.131	AS19871	US	31,238	3.82%
6	52.33.196.199	AS16509	US	23,994	2.93%
7	72.52.4.122	AS32787	US	21532	2.63%
8	93.46.8.89	AS12874	Italy	17,328	2.12%
9	72.52.4.119	AS32787	US	13,551	1.66%
10	118.193.172.49	AS58879	HK	10,689	1.31%
Total	-	-	-	313,373	38.32%

Table 3: Top 10 registrant emails.

No.	Email	Count of Levelsquatting e2LDs	Percentage
1	yu****@yinsibaohu.aliyun.com	3,328	5.19%
2	yuming****@163.com	2,985	4.66%
3	4645468b********@privacy.everdns.com	1,633	2.55%
4	zz****@sina.com	1,397	2.18%
5	28***@qq.com	1,255	1.96%
6	c138e837********@privacy.everdns.com	1,231	1.92%
7	xiaosh********@163.com	989	1.54%
8	ljj********@gmail.com	751	1.17%
9	whoisa****@west263.com	730	1.14%
10	zr**@qq.com	712	1.11%
Total	-	15,011	23.41%

Registration dates. Next, we examine the registration dates of the levelsquatting e2LDs. Figure 4 illustrates the ECDF of registration dates, which shows that more than 59.27% of domains were registered after 2016. Previous studies suggest recent registration date is an indicator of domains owned by attackers [21, 22], and our result suggests that hijacking reputable e2LD and adding subdomains under its zone file are not popular, since reputable e2LDs tend to have a long registration lifetime (*e.g.*, `google.com` has been registered for more than 20 years). Instead, creating e2LD or compromising newly registered e2LD is more popular.

Wildcard DNS. While `LDS` has detected 817,681 unique levelsquatting FQDNs, they are mapped to only 64,124 e2LDs. We suspect there may be many wildcard DNS records among them. To verify this assessment, we probe all 64,124 e2LDs using the same method proposed by Du *et al.* [15]. In essence, for an e2LD like `example.com`, we first try to resolve the IP address of `*.example.com`. The e2LD is considered to support wildcard DNS if there is a valid response. Otherwise, we issue two queries with random subdomain names, like `aaa.example.com` and `bbb.example.com`. If the two responses are matched, the e2LD is considered to support wildcard DNS as well. In the end, we discovered 41,389 e2LDs (64.55% of 64,124) contain wildcard DNS records, suggesting this configuration is widely used by adversaries.

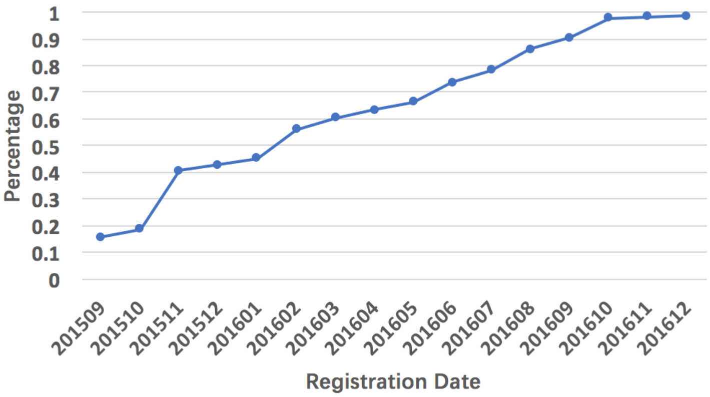


Fig. 4: ECDF of registration dates.

Table 4: Top 10 new gTLDs in levelsquatting e2LDs.

No.	New gTLD	Count	Percentage of new gTLD domains	Percentage of all e2LDs
1	.top	3,868	20.92%	6.03%
2	.win	3,034	16.41%	4.73%
3	.pw	2,672	14.45%	4.17%
4	.info	2,254	12.19%	3.52%
5	.bid	1,862	10.07%	2.90%
6	.loan	1,213	6.56%	1.89%
7	.party	1,021	5.52%	1.59%
8	.racing	893	4.83%	1.39%
9	.faith	586	3.17%	0.91%
10	.date	313	1.69%	0.49%
Total	-	17,716	95.83%	27.63%

Abuse of new gTLD domains. Previous studies [15] discovered that there is an increasing tendency of registering malicious domains under new gTLDs, like `.top`. We want to learn whether new gTLD is also favored by levelsquatting attackers. As such, we use the new gTLD list published by ICANN [23] to filter the e2LDs in DS_{All}. It turns out a prominent ratio of e2LDs (17,716, 27.63% of 64,124) are under new gTLDs, which aligns with the discovery of previous works. We think the the major reason is that most new gTLDs are cheap and lack of maintenance. We show the top 10 new gTLDs abused in Table 4.

SSL certificates. Deploying SSL certificates and supporting HTTPs connection is a growing trend for site administrators. To make malicious sites, especially phishing sites more convincing to visitors, SSL certificates are also used by attackers [24]. For levelsquatting domains, the motivation is the same but our measurement result shows that they have not seriously considered this option. We ran port scan with ZMap [24] over all DS_{All} and find that only 587 of them provide certificates. By comparison, a study [25] showed that already 70% of Alexa Top One Million sites provide SSL certificates. We

[24] `https://github.com/zmap/zmap`

download all these 587 certifications and extracted the issuers. Only **six** issuers are found. **All** of them can provide free SSL certification with 30-day period or even longer. We believe this is the main reason that these issuers are selected.

Table 5: SSL Certification issuers and domain count.

No.	Certification Issuer	Charge	Count	Percentage
1	RapidSSL SHA256 CA - G3	30 days free	276	47.02%
2	Let's Encrypt Authority X3	Free	207	35.26%
3	WoSign CA Free SSL Certificate G2	Free	40	6.81%
4	GlobalSign Organization Validation CA - G2	30 days free	26	4.43%
5	Cybertrust Japan Public CA G3	30 days free	23	3.92%
6	Amazon	12 month free	15	2.56%
Total	-	-	587	100%

6 Characterization

In this section, we take a closer look into the business behind levelsquatting domains and their targeted brands, to get a better understanding of how they serve attackers' operations.

6.1 Types of Malicious Activities

LDS is able to classify levelsquatting domains into two categories: phishing and non-phishing. In order to learn more finer-grained categorical information, *e.g.*, the business operated behind the domain, we extract more features from the associated pages and run another classification procedure. Specifically, we randomly sampled 10,000 pages from DS_{All} first and manually labeled them into 5 categories, including porn, lottery, phishing, blackhat SEO, malware-delivery to prepare the training dataset. Then, the texts from title and href tags of each page are extracted and we use a deep-learning algorithm, CNN (Convolutional Neural Network) to build the classification model [26]. We choose CNN because it has been applied to similar tasks like sentence and text classification, and achieved many successes [27, 28]. After the training step, we use CNN model to classify all DS_{All} pages. The result on levelsquatting FQDNs are shown in Figure 5. It turns out most of the levelsquatting domains were used for porn (42.59%) and lottery (34.42%).

Since the purposes of phishing sites are not always identical, we run the same CNN-based approach to obtain sub-categories under the phishing category. The statistics of the associated FQDNs are shown in Figure 6. It turns out the majority (94.89%) of FQDNs attempts to impersonate well-known sites of web portals, finance, advertisements and search-engine. Below we elaborate each category.

Fake web portals. The sites here are developed to help attackers gain high search rankings illicitly (*i.e.*, blackhat SEO). Attackers crawl content from reputable web portals and update everyday. Because the ranking algorithm favors sites with high dynamics and meaningful content, attackers' sites will gain relatively high score. In the mean time, blackhat SEO keywords and malicious URLs are embedded into the copied content. As a result, querying blackhat SEO keywords in search engines will lead to malicious URLs with higher possibilities [15].

No.	Type	Count	Percentage
1	Porn	348,233	42.59%
2	Lottery	281,425	34.42%
3	Phishing	137,388	16.80%
4	Blackhat SEO	40,316	4.93%
5	Malware delivery	2,893	0.35%
6	Others	7,426	0.91%
Total	-	817,681	100%

Fig. 5: Levelsquatting FQDN categories.

No.	Type	Count	Percentage
1	Fake web portal	45,783	33.32%
2	Fake finance	41,322	30.08%
3	Fake advertisement	29,925	21.78%
4	Fake search engine	13,331	9.70%
5	Fake domain Parking	1,937	1.41%
Total	-	132,298	96.30%

Fig. 6: Phishing FQDN sub-categories.

Fake financial sites. This is a classic type of phishing sites. Their goal is to steal users' credentials by impersonating the login pages of other sites, especially bank websites. These sites make themselves look almost the same as bank sites, stock buying and selling sites, to allure users to submit their bank card number and password.

Fake advertisements. These sites promote products by exaggerating their effects. For instance, fake weight-losing products are frequently seen. Their common strategy is to crawl the content from reputable shopping sites like `www.amazon.com` and replace some of the contents with fake advertisements.

Fake search engine. This is a *new* type of blackhat SEO that never reported before and we will show more details in Section 6.3. They pretend to be a valid search-engine site. A search query will be forwarded to the authentic site and the returned results would be mixed with illegal ads. As it is fully functional, users would prone to trust the returned results and click the illegal ads.

6.2 Visiting Strategies

A user could make a mistake when typing a domain name and visit a typo-squatting site accidentally, but it's not possible to type a levelsquatting domain name by mistake. So we wonder how these levelsquatting domains visited by users and who are their referrers.

Although it is straightforward to trace forward from a levelsquatting domain to its destination by following redirection and hyperlink, tracing *backward* is very challenging. A levelsquatting domain can also be embedded in webpage and many other media like email. Unfortunately, without data shared by service providers like email servers, finding the origin is impossible. We focus on websites that link to levelsquatting domains as we can leverage search engine, whose indexed pages are open to public, to find website origin. To this end, we queried all FQDNs in DS_{All} using Baidu and downloaded the first 3 result pages for each. We choose Baidu because Baidu allows us to run automated query without being blocked. In the end, we find only 80,159 queries returning at least one result, suggesting most of them are referred by other channels rather than websites.

The next step is to find pages in the search result that point to levelsquatting FQDNs. Instead of directly crawling, we choose to analyze its *short description* and look for FQDN in DS_{All}. To notice, short description of search result has been used for detecting promotional-infection in [20]. In the end, we found 298,370 search results satisfying this criteria. Interestingly, more than 27% of the results point to forums `zhidao.baidu.com` and `zhihu.com` (Chinese versions of Quora). We report these content to Baidu Security Team and all are removed now.

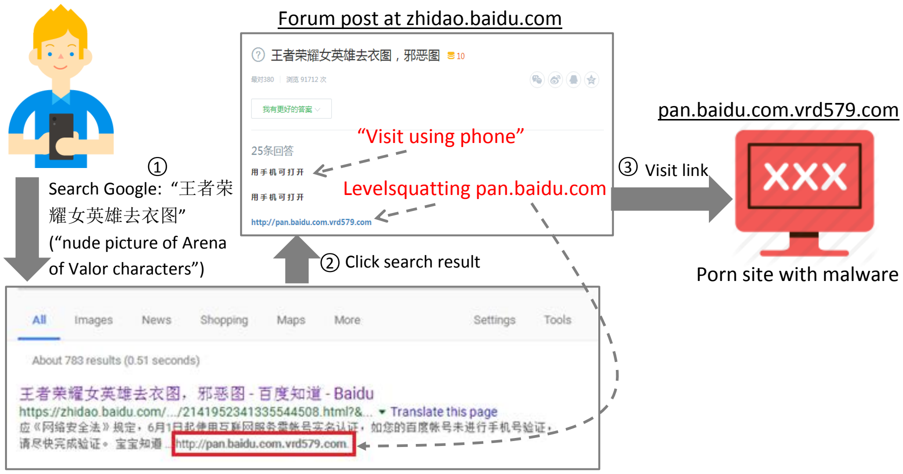

Forum post at zhidao.baidu.com
王者荣耀女英雄去衣图，邪恶图 10
最对380 | 浏览 91712 次
我有更好的答案
25条回答
用手机可打开
用手机可打开
http://pan.baidu.com.vrd579.com
"Visit using phone"
Levelsquatting pan.baidu.com
① Search Google: "王者荣耀女英雄去衣图" ("nude picture of Arena of Valor characters")
② Click search result
③ Visit link
pan.baidu.com.vrd579.com
XXX
Porn site with malware
All
Images
News
Shopping
Maps
More
Settings
Tools
About 783 results (0.51 seconds)
王者荣耀女英雄去衣图，邪恶图 - 百度知道 - Baidu
https://zhidao.baidu.com/.../2141952341335544508.html?&... ▾ Translate this page
应《网络安全法》规定，6月1日起使用互联网服务需帐号实名认证，如您的百度帐号未进行手机号验证，
请尽快完成验证。宝宝知道 ... http://pan.baidu.com.vrd579.com

Fig. 7: An example showing how a visitor reaches a levelsquatting domain.

Here we give a real-world example showing how a visitor reaches a levelsquatting domain from the website referral, and illustrate it in Figure 7. The attacker first posts a thread on `zhidao.baidu.com` which advertises a link pointing to `pan.baidu.com.vrd579.com`, a levelsquatting domain impersonating `pan.baidu.com`, Baidu's cloud-drive service. The thread tops the search result when a user queries "nude picture of Arena of Valor characters" (translated from Chinese). When user follows the search result and the link in the post by mobile, she will land on the levelsquatting site while only "pan.baidu.com" will be shown in her browser address bar, which will induce her to input her password or download malicious apps.

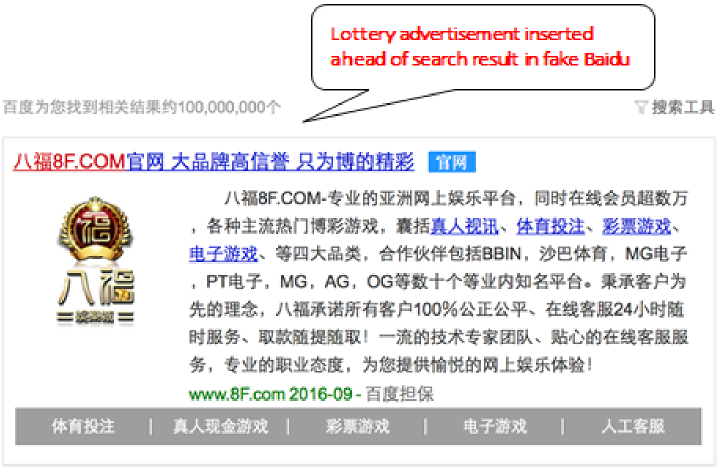
One may wonder if such search poisoning attack is only effective against Baidu. To examine this argument, we evaluated Google by sampling 10,000 domains in DS_{All} and querying them through our proxy pool. It turns out more than 85% levelsquatting domains were also indexed by Google.

6.3 An Example of Fake Search Engine

As pointed out in Section 6.1,during the course of our study, we have discovered a *new* type of phishing attack impersonating search engines. The fake search-engine site copies content from authentic site, but when a user searches a term, illegal ads are inserted ahead of the original search results. Figure 8 shows the returned page of `www.baidu.com.baidu-service.com` (impersonating Baidu search) when querying "abc." The first item is an advertisement pointing to a lottery site `8f.com`, which is not allowed in Baidu's search result because it's not permitted by the Chinese government. We count the number of levelsquatting domains under this category and find the top three are also the three leading search engines in China: Baidu, 360 search and Sogou. The fake sites count is 4,583, 3,950 and 2,318 seperately.

7 Browser UI Vulnerabilities

When the length of a domain name exceeds the visible area of browser's address bar, a part of the domain name will not be displayed. A user could mis-recognize the domain



Lottery advertisement inserted ahead of search result in fake Baidu

百度为您找到相关结果约100,000,000个

八福8F.COM官网 大品牌高信誉 只为博的精彩 官网

八福8F.COM-专业的亚洲网上娱乐平台，同时在线会员超数万，各种主流热门博彩游戏，囊括真人视讯、体育投注、彩票游戏、电子游戏、等四大品类，合作伙伴包括BBIN，沙巴体育，MG电子，PT电子，MG，AG，OG等数十个等业内知名平台。秉承客户为先的理念，八福承诺所有客户100%公正公平、在线客服24小时随时服务、取款随提随取！一流的技术专家团队、贴心的在线客服服务，专业的职业态度，为您提供愉悦的网上娱乐体验！

www.8F.com 2016-09 - 百度担保

Fig. 8: Fake Baidu search result.

in this case. Browser vendors should carefully design the address bar to either leave enough space for domain name or notify users when part of the domain, especially the e2LD section, is hidden. Unfortunately, not all browsers follow these design principles.

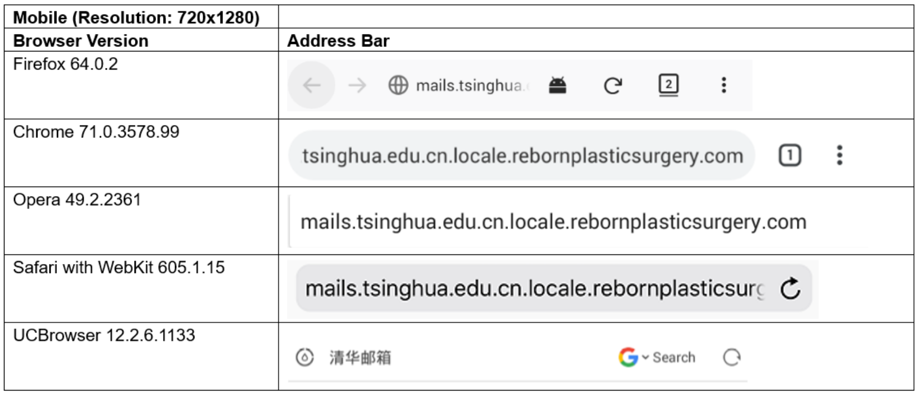
We first examine how a lengthy domain is displayed on mobile browsers. Five representative mobile browsers are tested through visiting `mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com` in an Android phone. Figure 9 shows the corresponding address bars. UC browser [25] is the most vulnerable as the domain name is entirely hidden in address bar. Similarly, Firefox only shows a few extra letters. Chrome and Safari perform better as more letters are displayed. We recommend Firefox and UC Browser to redesign the address bar for allowing better visibility. A work published recently [29] also pointed out that many mobile browsers fail to display levelsquatting domain name in a secured manner, which resonates with our findings. The desktop browsers are expected to be immune from this vulnerability, given that their UI has much larger visible area. We test 8 popular desktop browser and find only IE 9 partially displays as shown in Figure 10.

8 Discussion

Limitations. The criteria we enforce on the brand selection rule out potential levelsquatting domains that include typos (*e.g.*, `go0gle.com.example.com`) or overlap with only part of brand e2LDs (*e.g.*, `google.example.com`). The major reason is that finding true positives under these cases requires a lot more web crawling and queries against passive DNS. Besides, we argue that these domains are less likely to be created by attackers who have absolute freedom to fill the subdomain section with anything they like.

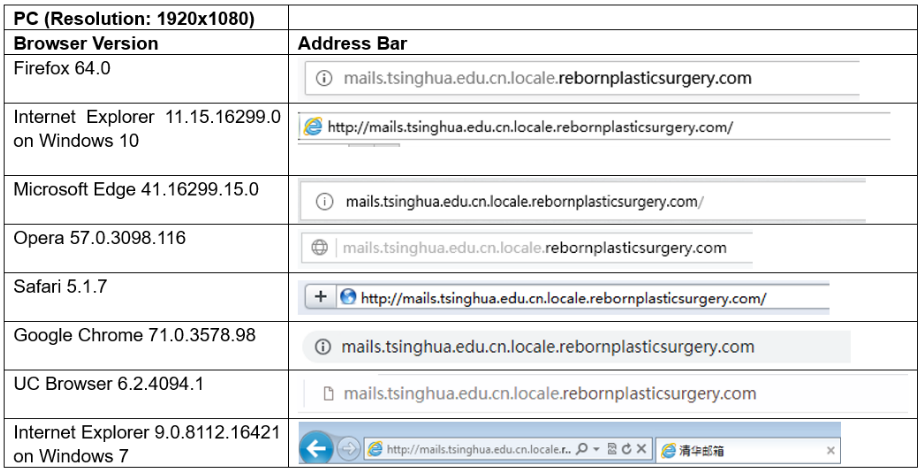
Knowing the design of `LDS`, attackers could adjust their strategies to avoid detection. For instance, they could target less popular brand domains (*i.e.*, beyond Alexa top 10K)

[25] `http://www.ucweb.com/`



Mobile (Resolution: 720x1280)	
Browser Version	**Address Bar**
Firefox 64.0.2	mails.tsinghua.
Chrome 71.0.3578.99	tsinghua.edu.cn.locale.rebornplasticsurgery.com
Opera 49.2.2361	mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com
Safari with WebKit 605.1.15	mails.tsinghua.edu.cn.locale.rebornplasticsurg
UCBrowser 12.2.6.1133	清华邮箱 Search

Fig. 9: Address bar of mobile browsers.



PC (Resolution: 1920x1080)	
Browser Version	**Address Bar**
Firefox 64.0	mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com
Internet Explorer 11.15.16299.0 on Windows 10	http://mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com/
Microsoft Edge 41.16299.15.0	mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com/
Opera 57.0.3098.116	mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com
Safari 5.1.7	http://mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com/
Google Chrome 71.0.3578.98	mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com
UC Browser 6.2.4094.1	mails.tsinghua.edu.cn.locale.rebornplasticsurgery.com
Internet Explorer 9.0.8112.16421 on Windows 7	http://mails.tsinghua.edu.cn.locale.r.. 清华邮箱

Fig. 10: Address bar of desktop browsers.

or change the page content to reduce the structure and visual similarity. These issues could be addressed when running other detection systems at the same time.

The majority of domain names inspected by `LDS` come from the passive DNS managed Qihoo 360. For this data, as far as we know, most of the logs are retrieved by DNS resolvers located in China. Thus, our measurement results could have certain bias towards one region, mainly about business categories and targeted brands (*e.g.*, Baidu has the most impersonators as shown in our result).

Suggestions to browser manufactures. We recommend browser companies to leave more space in the address bar. For example, a scroll bar could be activated when the domain overflows the display region to allow user to see the full name. Another way is to highlight the e2LD part in the address bar.

Suggestions to users. A domain name should be reviewed more carefully when it is lengthy or covers the entire region of the address bar. The entire domain name should be inspected, not only the beginning section. If the e2LD section is suspicious or never seen before, the domain should be avoided.

Suggestions to registrars. We suggest registrars to adjust their policy to limit the length and depth of a subdomain, given that normal domains rarely have so many characters or levels. Alternatively, registrars can enforce a rule to forbid a domain owner to create a subdomain with multiple levels at one shot.

Responsible disclosure. We have reported 4,583 fake search-engine sites impersonating Baidu search and 38,275 pages embedding `baidu.com` in the subdomain section to Baidu security team. All of them have been confirmed malicious. In addition, the posts under `zhidao.baidu.com` backlinked by the malicious pages are all removed.

Regarding the browser UI vulnerabilities, we have contacted several browser vendors including Baidu browser [26] and 360 security browser [27] which have the similar issue as UC browser. Their security teams have acknowledged our findings and have fixed in the current browsers.

9 Related Work

Domain-squatting. Various domain-squatting attacks have been discovered and studied before, including typo-squatting [6], bit-squatting [7], homophone-squatting [8], homograph-squatting [9], combosquatting [10] and etc. Attackers under these scenarios all need to *buy* an e2LD and create domains *similar* to the brand domain. They can be thwarted when the owner actively registers adjacent domains or by detection mechanisms based on domain-name similarity [6,30–32]. However, such methods fail to defend from levelsquatting attack since the subdomain can be arbitrarily created by attackers under any e2LD not controlled by brand owners.

Domain abuse. Understanding how attackers register and use domains is essential for detecting malicious domains. Previous works have extensively studied attackers' strategies and patterns in domain registration [34–36]. Their studies show attackers' preferences of registrars with loose regulations. In [29], authors focused UI vulnerabilities in mobile browsers and gave a systematic measurement on security vulnerabilities in them.

Our work leverages passive DNS data, registration data and visual similarity to discover levelsquatting domains. Passive DNS data has been extensively leveraged for detecting botnet and spam domains [37–40]. Recently, the data from domain registrars has shown potential in detecting domain abuse at the early stage [21,41]. We leverage similarity-based approach to detect phishing levelsquatting domains, which aligns with previous works in this area [13,42,43].

We compared our work with [44], which studied how adversaries use subdomains created under the *compromised* e2LDs for malign purposes. In our work, we focus on the attack that utilizes subdomains created to impersonate reputable domains. Attackers only need to buy a cheap domain name and create reputable prefix purposefully, and they do not need to compromise legitimate domains.

Underground economy. Our study shows that levelsquatting is extensively used by the underground economy to deceive web users. This "dark" community has been investigated by many researchers in order to gain better understanding about its operational

[26] `https://liulanqi.baidu.com/`

[27] `http://se.360.cn/`

model and build effective defense. On this topic, Levchenko et al. [16] revealed the infrastructure and strategies used for email spam. Nektarios et al. [45] studied how illicit drug trade was facilitated through search-redirection attack. Many of the underground transactions happen at anonymous marketplace. Its scale and operational model were studied by Nicolas *et al.* [46], and Barratt *et al.* [47].

10 Conclusion

In this work, we present a study about the phenomenon of levelsquatting, which exploits visual vulnerabilities of browsers to defraud web users. In order to obtain sufficient amount of data, we have developed a system named `LDS`, which examines a large volume of passive DNS data and applies three different checkers to detect levelsquatting domains. In the end we have identified 817,681 malicious FQDNs with an accuracy of 96.9%.

Based on the data produced by `LDS` and obtained from VirusTotal and PhishTank, we carried out a comprehensive study to understand the impact of this threat and the strategies used by attackers. Our study has already revealed several unique insights, like prefixes favored by attackers. We also discovered a new type of phishing attack against search-engine. Furthermore, we analyze how levelsquatting domain name is displayed in mobile and desktop browsers and find 3 browsers (2 mobile and 1 desktop) display domain names in a misleading way. We have reported our findings to Baidu security team and 360 security, receiving very positive feedback.

Our study shows that attackers are constantly exploiting the weakness of domain ecosystem and inventing new attack vectors. In the future, we will continue the research regarding domain abuse with a focus on its impact and the new trend.

11 Acknowledgments

We thank anonymous reviewers for their insightful comments. This work is supported in part by the National Natural Science Foundation of China (U1836213, U1636204) and the BNRist Network and Software Security Research Program (Grant No. BN-R2019TD01004).

References

1. What Is TLDR? https://www.lifewire.com/what-is-tldr-2483633. 2017
2. How scammers use sub-domains. http://easykey.uk/computer-safety/how-scammers-use-sub-domains. 2016
3. Yang, Hao, et al. "How to learn klingon without a dictionary: Detection and measurement of black keywords used by the underground economy." 2017 IEEE Symposium on Security and Privacy (SP). IEEE, 2017.
4. Marchal, Samuel, Jérôme François, and Thomas Engel. "Proactive discovery of phishing related domain names." International Workshop on Recent Advances in Intrusion Detection. Springer, Berlin, Heidelberg, 2012.
5. DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION. https://tools.ietf.org/html/rfc1035. 1987.
6. Wang, Yi-Min, et al. "Strider Typo-Patrol: Discovery and Analysis of Systematic Typo-Squatting." SRUTI 6.31-36 (2006): 2-2.

7. Nikiforakis, Nick, et al. "Bitsquatting: Exploiting bit-flips for fun, or profit?." Proceedings of the 22nd international conference on World Wide Web. ACM, 2013.
8. Wiener, Seth. "Grass-mud horses to victory: The phonological constraints of subversive puns." Proceedings of the 23rd North American Conference on Chinese Linguistics. Vol. 1. 2011.
9. Holgers, Tobias, David E. Watson, and Steven D. Gribble. "Cutting through the Confusion: A Measurement Study of Homograph Attacks." USENIX Annual Technical Conference, General Track. 2006.
10. Kintis, Panagiotis, et al. "Hiding in plain sight: A longitudinal study of combosquatting abuse." Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017.
11. Phishing with 'punycode' – when foreign letters spell English words. https://nakedsecurity.sophos.com/2017/04/19/phishing-with-punycode-when-foreign-letters-spell-english-words/. 2017
12. Liu, Baojun, et al. "A reexamination of internationalized domain names: the good, the bad and the ugly." 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. 2018.
13. Maurer, Max-Emanuel, and Dennis Herzner. "Using visual website similarity for phishing detection and reporting." CHI'12 extended abstracts on human factors in computing systems. ACM, 2012.
14. Levchenko K, Pitsillidis A, Chachra N, et al. Click trajectories: End-to-end analysis of the spam value chain[C]//2011 ieee symposium on security and privacy. IEEE, 2011: 431-446.
15. Du, Kun, et al. "The Ever-Changing Labyrinth: A Large-Scale Analysis of Wildcard DNS Powered Blackhat SEO." 25th USENIX Security Symposium (USENIX Security 16). 2016.
16. Levchenko, Kirill, et al. "Click trajectories: End-to-end analysis of the spam value chain." 2011 ieee symposium on security and privacy. IEEE, 2011.
17. Wang, David Y., Stefan Savage, and Geoffrey M. Voelker. "Cloak and dagger: dynamics of web search cloaking." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.
18. Zhang, Qing, David Y. Wang, and Geoffrey M. Voelker. "DSpin: Detecting Automatically Spun Content on the Web." NDSS. 2014.
19. Corona I, Biggio B, Contini M, et al. Deltaphish: Detecting phishing webpages in compromised websites[C]//European Symposium on Research in Computer Security. Springer, Cham, 2017: 370-388.
20. Liao, Xiaojing, et al. "Seeking nonsense, looking for trouble: Efficient promotional-infection detection through semantic inconsistency search." 2016 IEEE Symposium on Security and Privacy (SP). IEEE, 2016.
21. Hao, Shuang, et al. "PREDATOR: proactive recognition and elimination of domain abuse at time-of-registration." Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016.
22. Li, Zhou, et al. "Knowing your enemy: understanding and detecting malicious web advertising." Proceedings of the 2012 ACM conference on Computer and communications security. ACM, 2012.
23. new gTLD Statistics by Top-Level Domains. https://ntldstats.com/tld. 2016
24. Nagunwa, Thomas. "Behind identity theft and fraud in cyberspace: the current landscape of phishing vectors." International Journal of Cyber-Security and Digital Forensics (IJCSDF) 3.1 (2014): 72-83.
25. TLS Certificates from the Top Million Sites. https://adamcaudill.com/2016/09/23/tls-certificates-top-million-sites/. 2016
26. Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.

27. Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom. "A convolutional neural network for modelling sentences." arXiv preprint arXiv:1404.2188 (2014).
28. Liu, Pengfei, Xipeng Qiu, and Xuanjing Huang. "Recurrent neural network for text classification with multi-task learning." arXiv preprint arXiv:1605.05101 (2016).
29. Luo, Meng, et al. "Hindsight: Understanding the evolution of ui vulnerabilities in mobile browsers." Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017.
30. Chen, Guanchen, et al. "Combating typo-squatting for safer browsing." 2009 International Conference on Advanced Information Networking and Applications Workshops. IEEE, 2009.
31. Banerjee, Anirban, Md Sazzadur Rahman, and Michalis Faloutsos. "SUT: Quantifying and mitigating url typosquatting." Computer Networks 55.13 (2011): 3001-3014.
32. Linari, Alessandro, et al. "Typo-Squatting: The Curse"of Popularity." (2009).
33. Agten, Pieter, et al. "Seven Months' Worth of Mistakes: A Longitudinal Study of Typosquatting Abuse." NDSS. 2015.
34. Hao, Shuang, Nick Feamster, and Ramakant Pandrangi. "Monitoring the initial DNS behavior of malicious domains." Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.
35. Coull, Scott E., et al. "Understanding domain registration abuses." Computers & security 31.7 (2012): 806-815.
36. Anderson, David S., et al. Spamscatter: Characterizing internet scam hosting infrastructure. Diss. University of California, San Diego, 2007.
37. Antonakakis, Manos, et al. "Building a dynamic reputation system for dns." USENIX security symposium. 2010.
38. Antonakakis, Manos, et al. "From throw-away traffic to bots: detecting the rise of DGA-based malware." Presented as part of the 21st USENIX Security Symposium (USENIX Security 12). 2012.
39. Bilge, Leyla, et al. "EXPOSURE: Finding Malicious Domains Using Passive DNS Analysis." Ndss. 2011.
40. Antonakakis, Manos, et al. "Detecting Malware Domains at the Upper DNS Hierarchy." USENIX security symposium. Vol. 11. 2011.
41. Lever, Chaz, et al. "Domain-Z: 28 registrations later measuring the exploitation of residual trust in domains." 2016 IEEE Symposium on Security and Privacy (SP). IEEE, 2016.
42. Garera, Sujata, et al. "A framework for detection and measurement of phishing attacks." Proceedings of the 2007 ACM workshop on Recurring malcode. ACM, 2007.
43. Medvet, Eric, Engin Kirda, and Christopher Kruegel. "Visual-similarity-based phishing detection." Proceedings of the 4th international conference on Security and privacy in communication netowrks. ACM, 2008.
44. Liu, Daiping, et al. "Don't Let One Rotten Apple Spoil the Whole Barrel: Towards Automated Detection of Shadowed Domains." Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2017.
45. Leontiadis, Nektarios, Tyler Moore, and Nicolas Christin. "Measuring and Analyzing Search-Redirection Attacks in the Illicit Online Prescription Drug Trade." USENIX Security Symposium. Vol. 11. 2011.
46. Christin, Nicolas. "Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace." Proceedings of the 22nd international conference on World Wide Web. ACM, 2013.
47. Barratt, Monica J., Jason A. Ferris, and Adam R. Winstock. "Use of S ilk R oad, the online drug marketplace, in the United Kingdom, A ustralia and the United States." Addiction 109.5 (2014): 774-783.